Supplement to

Long vs. Short: Understanding the dynamics of persistent summer hot spells in Europe

Duncan Pappert et al.

Correspondence to: Duncan Pappert (duncan.pappert@unibe.ch)

List of material:

This document contains relevant material that is additional to the main manuscript, structured in five parts. Sections 1 and 2 add context to some of the methodological choices made in the study regarding the regionalisation and hot spell computation, respectively. Section 3 includes additional figures related to SW and W Europe, which complement the analysis in the main article and are sometimes referenced within it. Section 4 illustrates findings regarding the other four regions not covered in the study. Finally, Section 5 shows the composited long spells across their whole lifetime, with no subsampling. The last three chapters contain only the figures with captions and no accompanying text.

1 More on the regionalisation

The main manuscript introduces a regionalisation to cluster locations in Europe based on their tendency to experience concurrent near-surface temperature extremes. Three-week non-overlapping averages of summer 2-meter temperature (T2M) time series of gridcells over land were binarised based on their exceedance of their 95^{th} summer percentile calculated over the full period (1959-2022). Gridcells were then clustered into distinctive regions based on a hierarchical agglomerative clustering using the Jaccard distance metric applied to the binary time series. Truncating the dendrogram at the desired level of event co-occurrence (12.5%) yielded the six regions presented in the study. Figure S1 shows the 'true' footprint of these regions, i.e. the average intensity and extent of the hot spells experienced by each area.

There was no real need to seek an objectively optimal number of clusters. Most regional studies of European heatwaves reference in our study focus on six regions (see Sect. 3.1 of main manuscript) and these more or less always tend to be the same ones. The results of our regionalisation are naturally influenced by the choice of parameters. Figure S2 shows that, keeping the truncation and temperature extremeness unchanged, a shorter temporal aggregation of T2M anomalies results in more and smaller regions. The same applies when increasing the extreme threshold for the series binarisation (Fig. S3 or choosing a lower truncation distance. There are many possible combinations and these can be adjusted depending on the specific impacts that are of interest.

Furthermore, several combinations of these parameters can result in six regions (see Table S1) that resemble each other to some degree. In other words, such is the extreme temperature variability in Europe that the clustering will inevitably always reflect the regional climates of the continent, i.e. dry and humid subtropical, temperate oceanic/maritime, continental, and subarctic climates. Almost all six regions from the different configurations will fit into these categories. Figure S4 expresses this point by showing the overlap of the nine clustering combinations. There typically tends to be a cluster for the Iberian Peninsula, sometimes extending to Southern France or Northern Italy. The Western European cluster may or may not include the British Isles and/or Southern Scandinavia. The regions covered by Northern clusters 4 and 6 display the most variability in spatial frequency.

The spatial robustness of these regions and similarities with the literature justify settling on six clusters. The higher the number of clusters, the more variability and inhomogeneities likely to arise between different clustering set ups. In this context, the reasoning behind parameter selection becomes increasingly important.

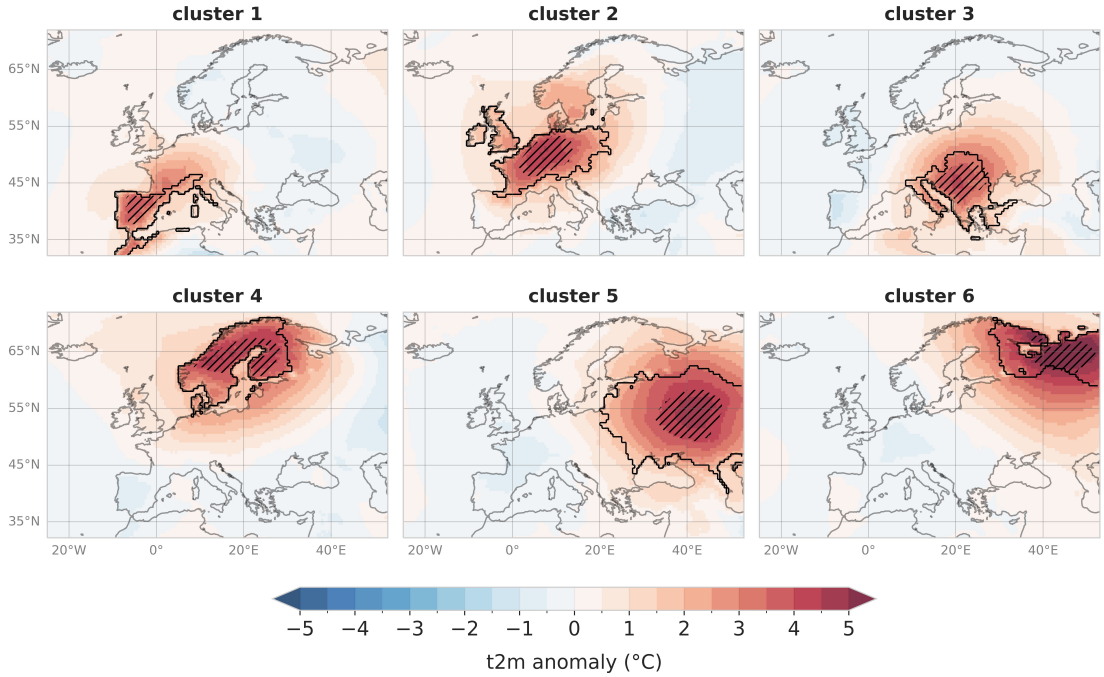

Figure S1. 'True' footprint of the six clusters obtained from the regionalisation in the study. Filled contours show the composited T2M anomalies of all identified hot spells. Hatched areas denote gridcells that have an 80% probability of exceeding the cluster-averaged 1 standard deviation (i.e. the threshold for defining a hot spell day).

Table S1. Nine combinations of clustering parameters that yield six final regions; i* is the setup used in our study.

#	temporal aggregation (days)	extreme threshold (%ile)	truncation distance (d_{trunc})	final number of regions
i*	21	95	.875	6
ii	21	90	.825	6
iii	21	85	.785	6
iv	14	95	.875	6
v	14	90	.825	6
vi	14	85	.775	6
vii	7	95	.900	6
viii	7	90	.815	6
ix	7	85	.790	6

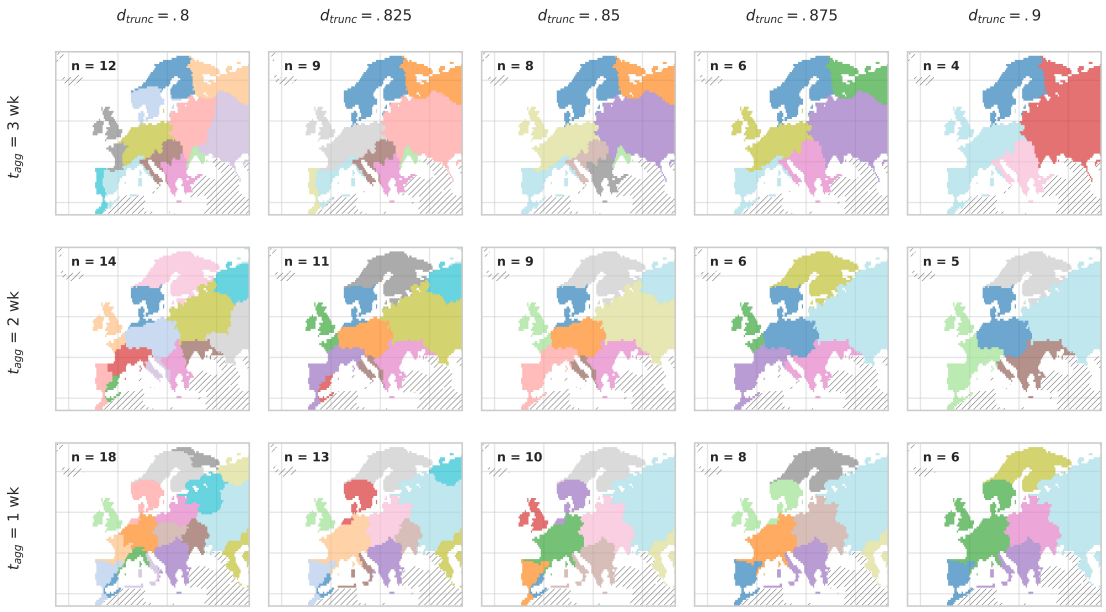

Figure S2. Outcome of regionalisation for different parameter combinations with a 95th percentile threshold exceedance. Rows show different time aggregations of 3, 2 and 1 week. Columns represent different final truncation distances of the clustering dendrogram. The 'n' in each panel is the number of clusters. The second panel from the right in the top-most row shows the 6 regions we use in our study.

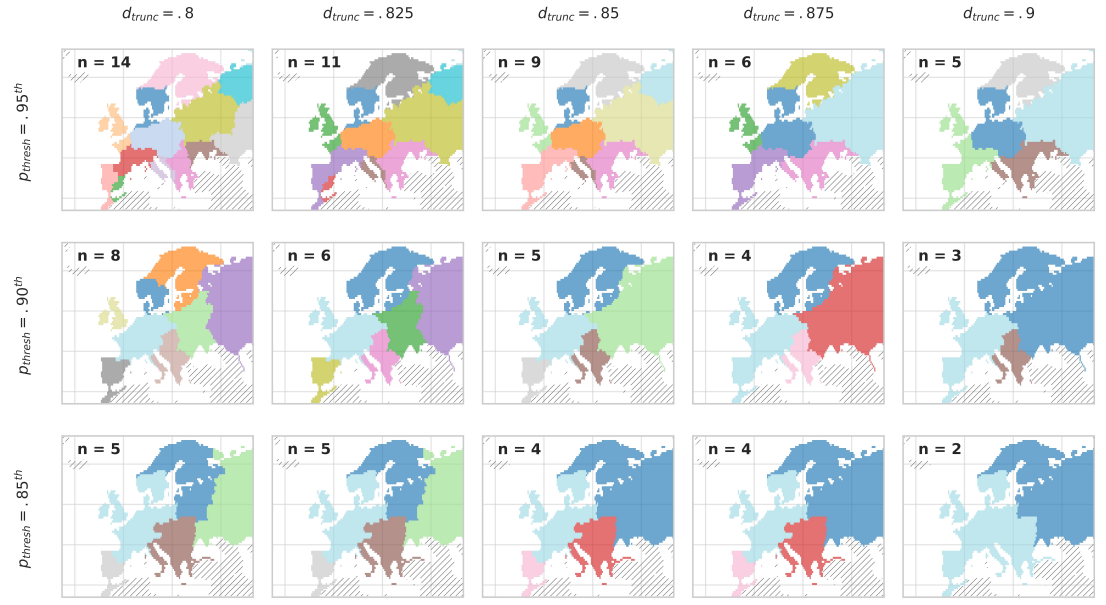

Figure S3. Outcome of regionalisation for different parameter combinations with a 2-week temporal aggregation. Rows show different extreme threshold exceedances 95^{th}, 90^{th}, and 85^{th} percentiles. Columns represent different final truncation distances of the clustering dendrogram. The 'n' in each panel is the number of clusters.

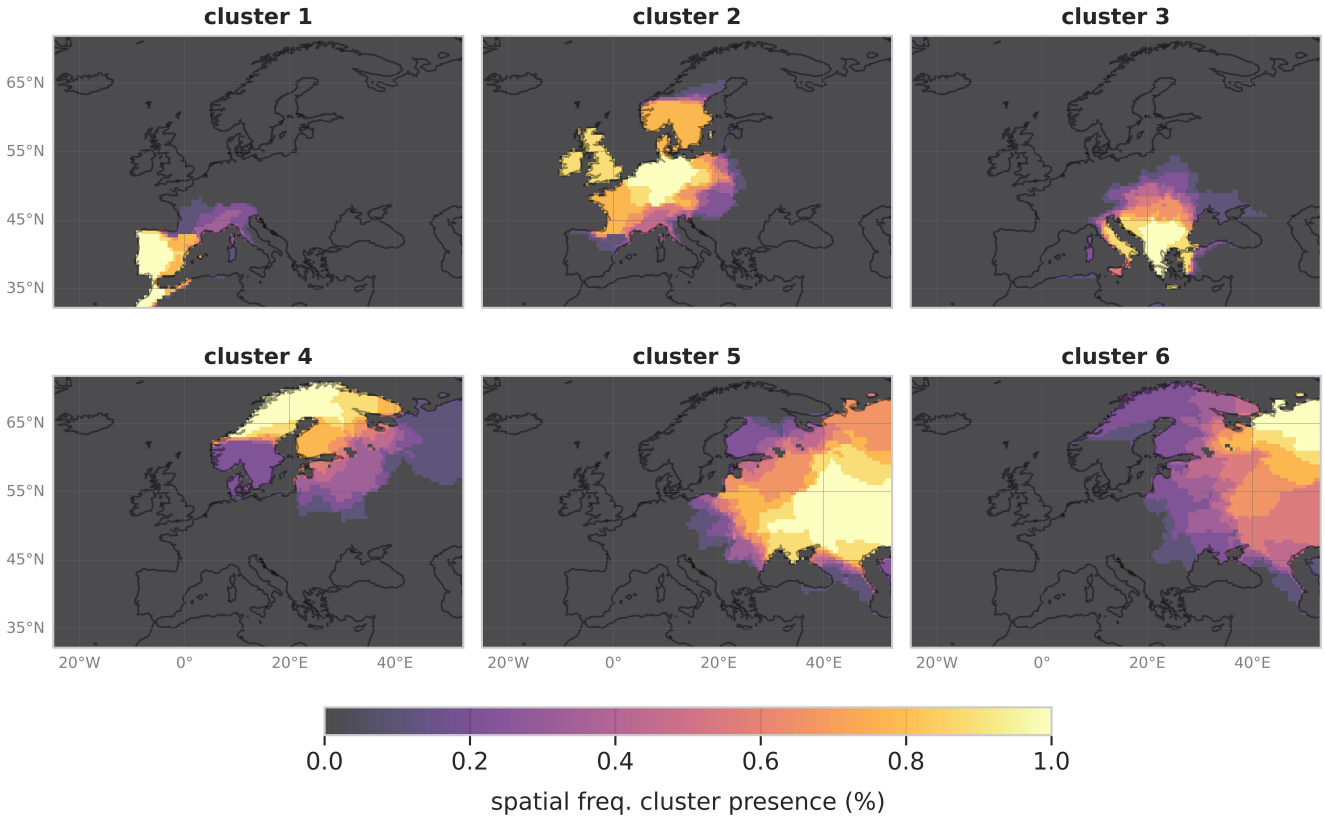

Figure S4. Spatial frequency of six regions for the nine different clustering outcomes, based on parameter combinations from Table S1.

2 Multiple hot spell definitions

Once the study regions are obtained, there a number of possible ways to compute hot spells. While there is no single agreed-upon (or even correct) way to define them, different methods will result in different events, sample sizes, and statistics. In our study, we computed the field mean of standardised T2M anomalies to obtain cluster-averaged series, from which distribution we then calculated the 'hot' day exceedance threshold (Fig. S5a). Another option would have been simply to look if the daily cluster-average of standardised anomalies exceeds a specified threshold (Fig. S5b). Alternatively, rather than working with spatial means, one could instead require a certain fraction of gridcells in the region to exceed the threshold on a given day (Fig. S5c). Other choices have to be made to compute hot spells, such as define the extremeness of the threshold (1σ in the following examples), as well as the number of non-hot days allowed to merge hot spells (set to 2 in the following examples), both of which define and affect the types of events one is dealing with.

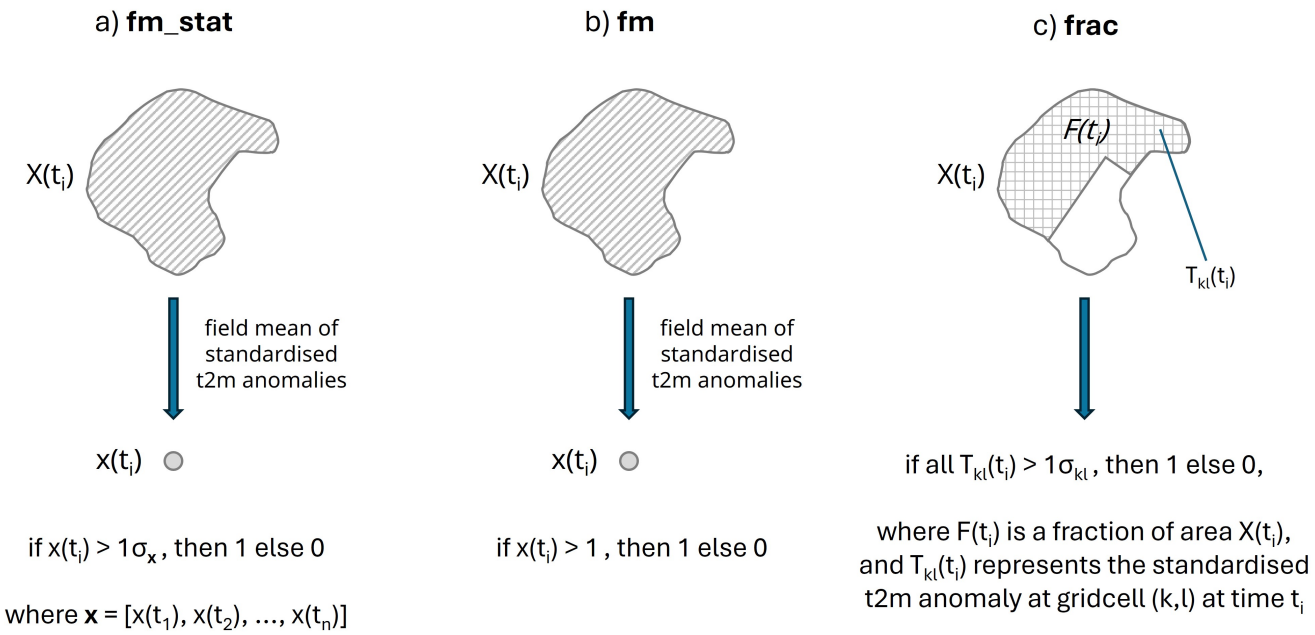

Figure S5. Schematic illustration of three different methods to compute hot spells, using threshold exceedances of a standardised T2M anomaly field. These are labelled (a) 'fm_stat' to refer to the statistical distribution of the field mean series, (b) 'fm' for the simple field mean, and (c) 'frac' to denote the fractional requirement for a threshold exceedance.

Figure S6 illustrates how these different hot spell computation methods affect the number and types of events that are identified in the six study regions. The exact number of short (4-5d), long (12-26d), and total hot spells is reported in Table S2. It is apparent that the method we employed in our study is the least stringent and yields the highest number of persistent hot spells (fatter right tail). It has a similar distribution to the method requiring that at least 33% of gridcells in the regions exceed their respective 1σ. Comparatively, the 'fm' and 'frac (F=0.66)'

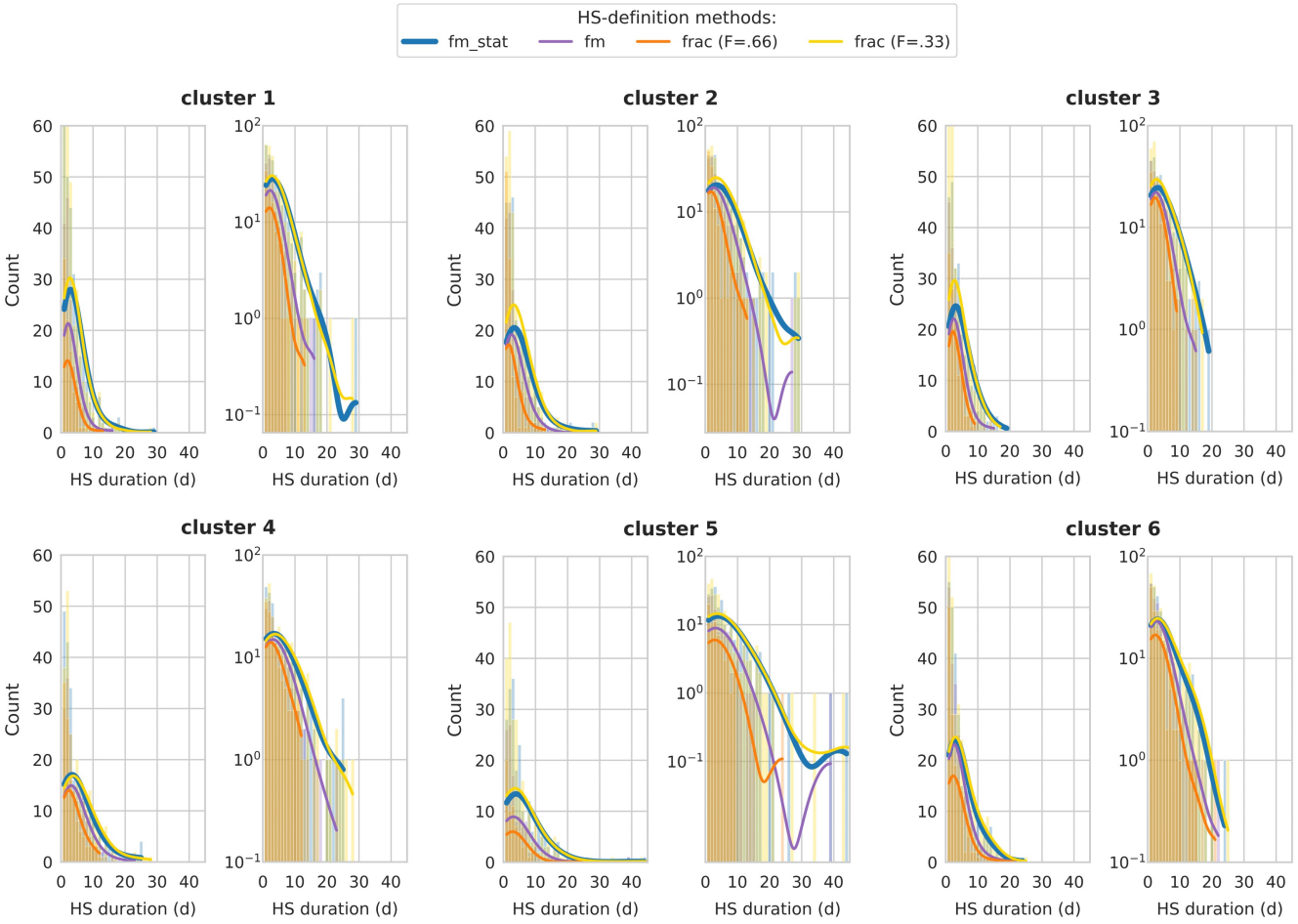

Figure S6. Histograms displaying hot spell duration counts for each of the six regions in our study according to different HS definition *fm_stat*, *fm*, *frac* (with F=.66), *frac* (with F=.33). Shown for each region subplot are the counts in arithmetic scale (left panels) and logarithmic scale (right panels).

approaches result in fewer events overall. These events are on average more extreme and over a larger area of the regions, as is visible in the example for W Europe in Figure S7, followed by 'fm_stat', and finally 'frac (F=0.33)' with the weakest anomalies. Conversely, 'fm_stat' has the lowest T2M variance over the Euro-Atlantic region of all methods (Fig. S8). This is likely due to the fact that 'fm_stat' considers the distribution of the cluster-averaged time series, meaning it is less influenced by outliers or unusual temperature spikes.

Table S2. Number of hot spells (HS) in all six clusters/regions resulting from different HS definitions. The threshold for each method is $+1\sigma$. F denotes the fraction of gridcells of the total cluster area used to calculate the threshold exceedances for method 'frac'. Exceedances separated by up to 2 days were merged.

events cluster 1	fm_stat	fm	frac (F=.66)	frac (F=.33)
4-5d HS	57	22	14	49
12-26d HS	16	3	1	15
all HS	280	153	91	292
events cluster 2				
4-5d HS	39	28	8	45
12-26d HS	15	3	2	16
all HS	258	178	119	298
events cluster 3				
4-5d HS	52	34	27	44
12-26d HS	13	4	0	11
all HS	255	170	121	288
events cluster 4				
4-5d HS	33	27	21	34
12-26d HS	26	9	1	27
all HS	242	168	125	247
events cluster 5				
4-5d HS	41	17	13	41
12-26d HS	24	3	1	24
all HS	210	115	65	232
events cluster 6				
4-5d HS	44	38	30	55
12-26d HS	24	6	4	31
all HS	272	206	146	296

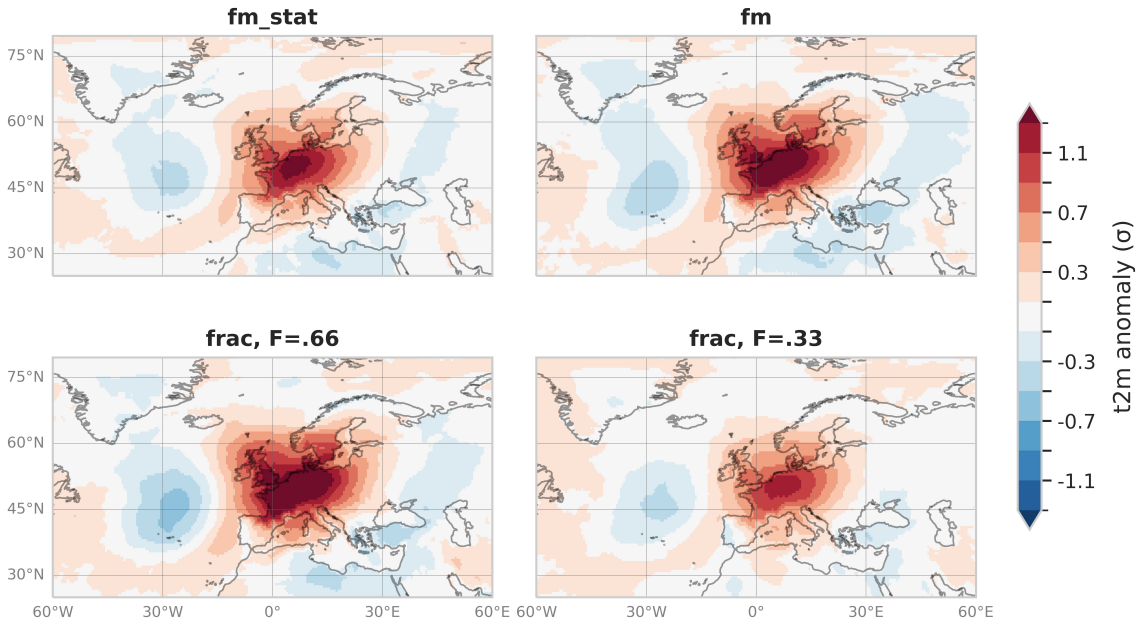

Figure S7. Standardised T2M anomaly composites for hot spells detected in W Europe (region 2) with different identification methods. Hot spells were randomly subsampled to match the (minimum) total number of events across methods, i.e. in our case *frac* (F=.66), with n=119 (see Table S2).

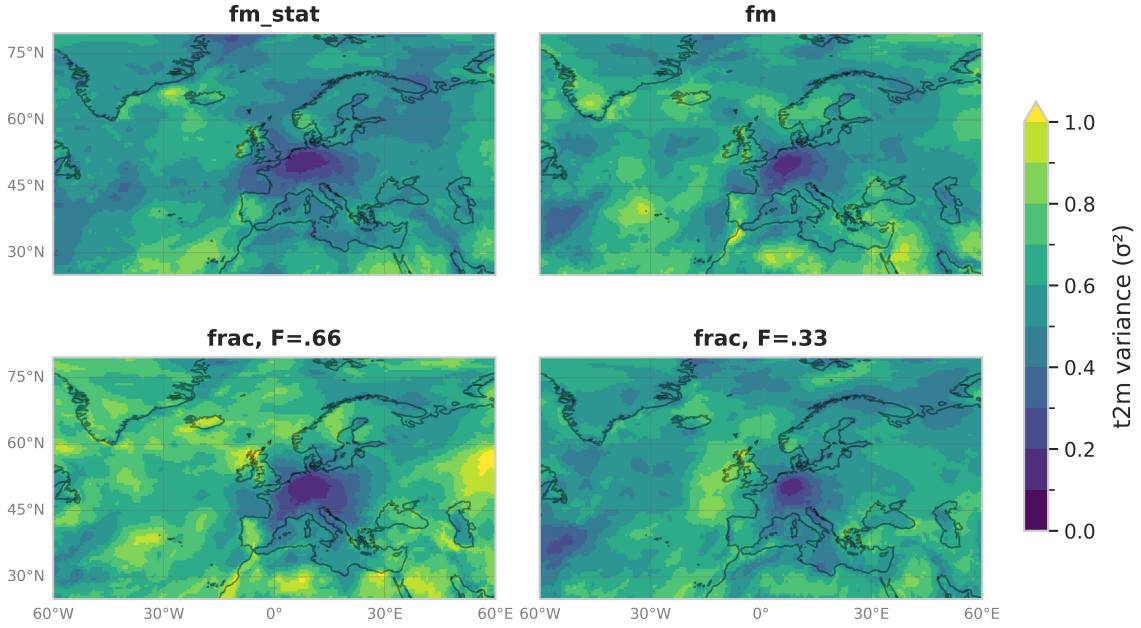

Figure S8. Variance composites of standardised T2M anomalies for hot spells detected in W Europe (region 2) with different identification methods. Hot spells were randomly subsampled to match the (minimum) total number of events across methods, i.e. in our case *frac* (F=.66), with n=119 (see Table S2).

3 Additional figures for SW and W Europe (regions 1 & 2)

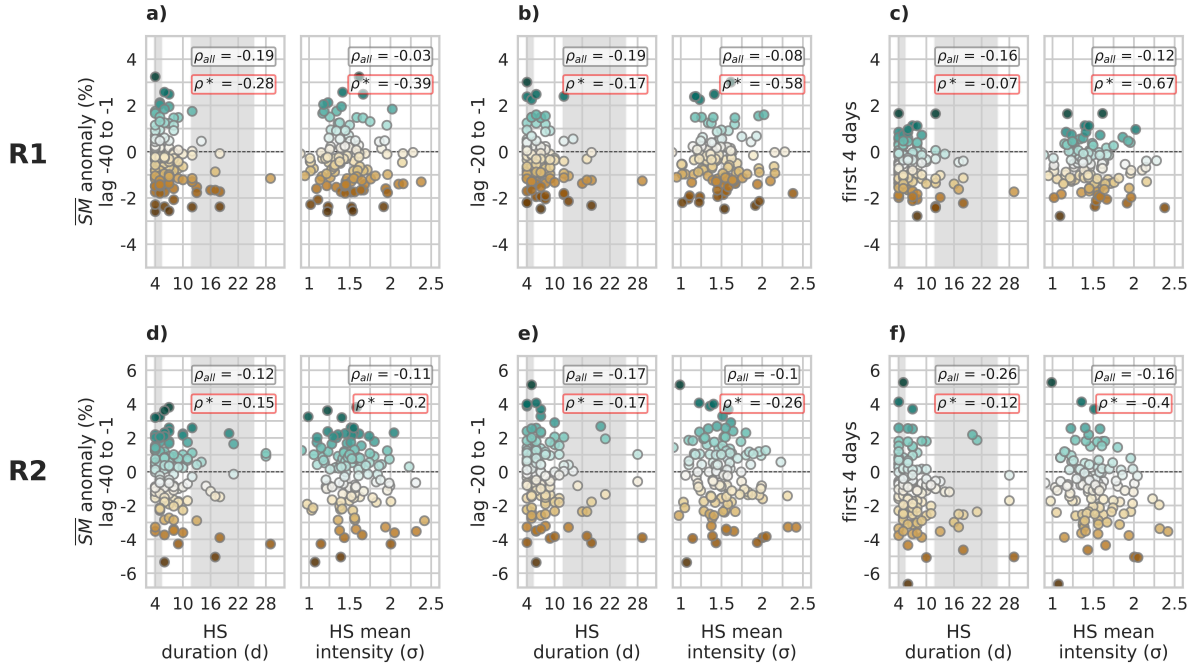

Figure S9. Mean soil moisture anomalies at different lags around the hot spells plotted against hot spell duration and mean intensity for SW Europe (R1) and W Europe (R2). ρ_{all} and $\rho*$ denote the Spearman correlation coefficients for all hot spells longer than 4 days and for just the long hot spells as defined in the study (lasting 12 to 26 days, i.e. the points falling within the wider gray shading of the duration subplots).

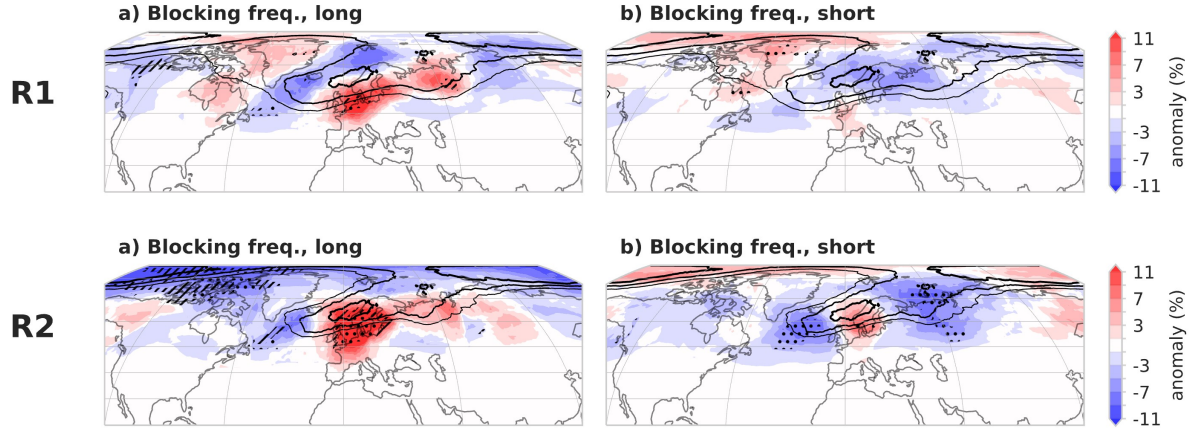

Figure S10. Spatial composites of blocking frequency anomalies for long (left) and short (right) hot spells for SW Europe (R1, top) and W Europe (R2, bottom). The blocking events are detected as positive Z500 anomalies, in contrast to the VIPV-based outcome described in Sect. 3.3 of the main article.

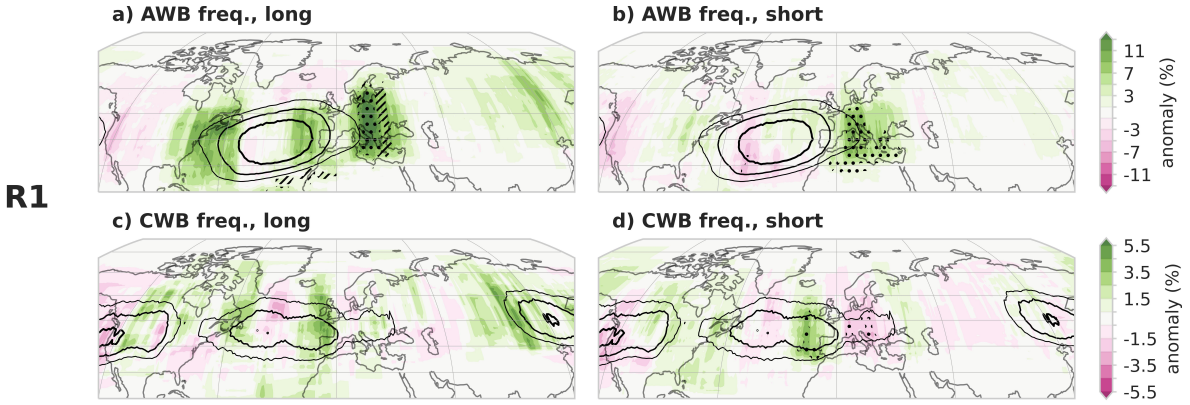

Figure S11. Composite fields of overturning (a,b) anticyclonic and (c,d) cyclonic Rossby wavebreaking frequency anomalies during long and short hot spells in SW Europe (R1).

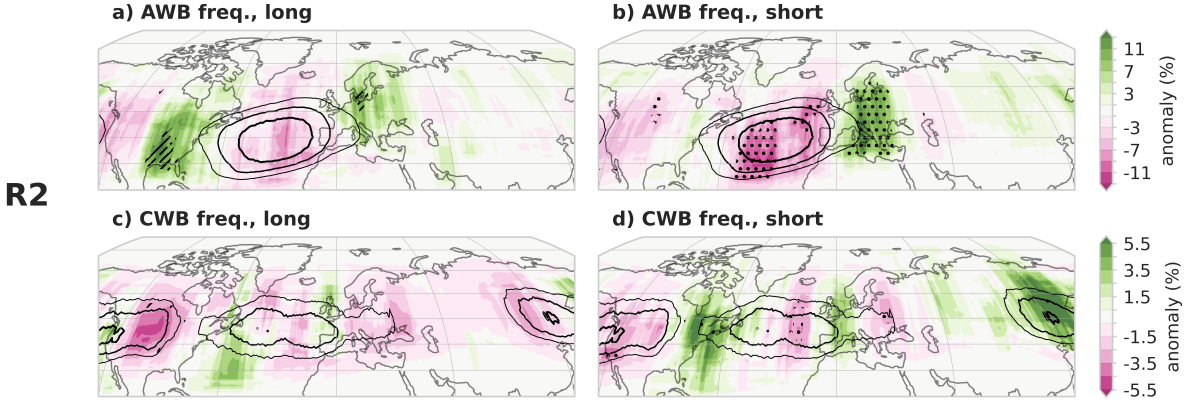

Figure S12. Composite fields of overturning (a,b) anticyclonic and (c,d) cyclonic Rossby wavebreaking frequency anomalies during long and short hot spells in W Europe (R2).

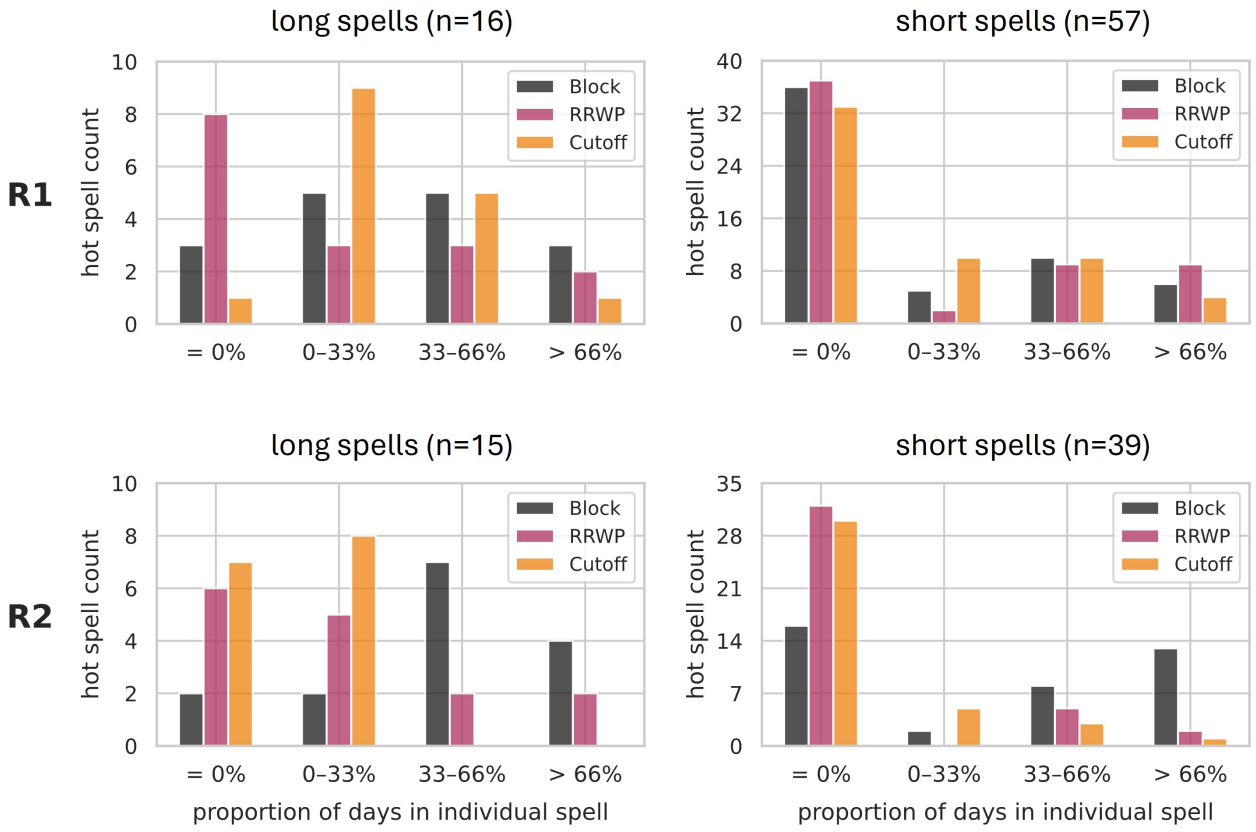

Figure S13. Relative presence of blocking, RRWPs and cutoffs during long (left) and short (right) hot spells in SW Europe (R1, top) and W Europe (R2, bottom). The four groups in the x axis denote absence of a driver (=0%), at least one-third presence (0-33%), between one- and two-thirds presence (33-66%), and more than two-thirds presence (>66%) during hot spell lifetime.

4 Results for regions 3, 4, 5, and 6

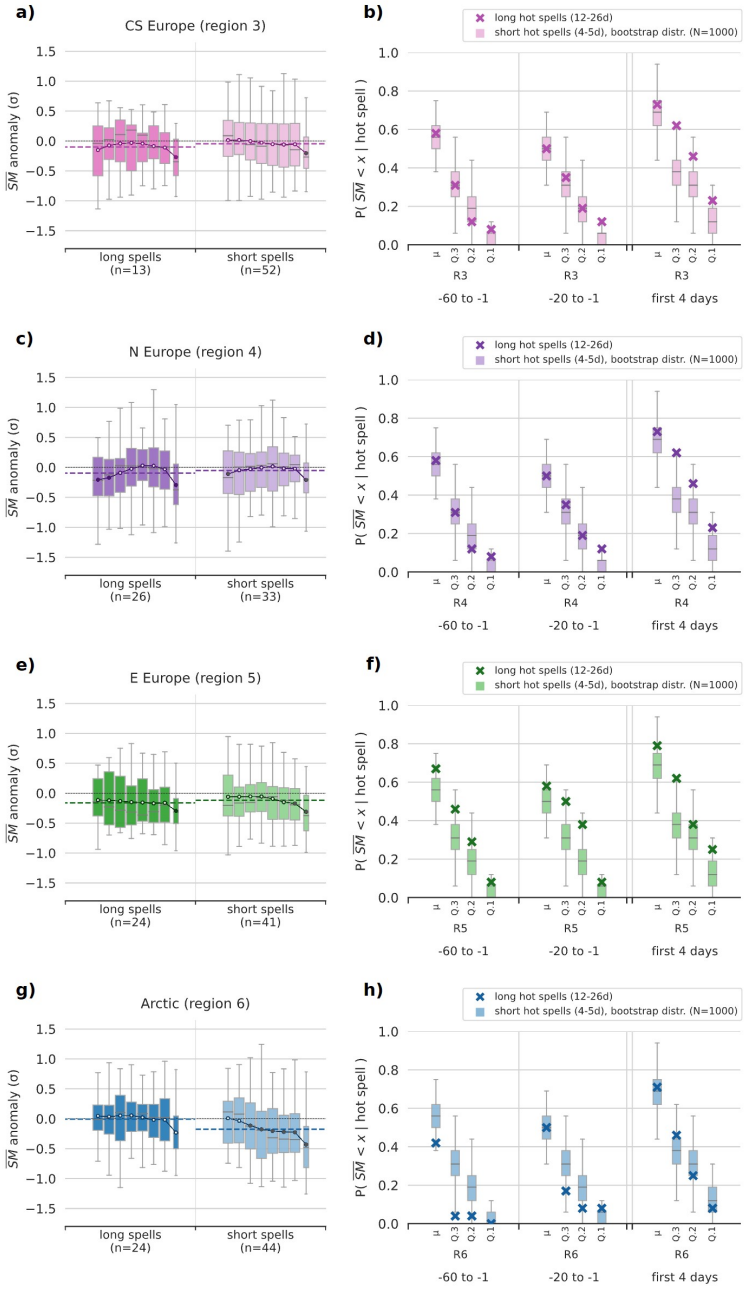

Figure S14. SM analysis analogous to Fig. 3 in Sect. 3.2 from the main manuscript for CS Europe (region 3, a-b), N Europe (region 4, c-d), E Europe (region 5, e-f) and Arctic/NW Russia (region 6, g-h). Panels on the left side (a,c,e,g) show region-averaged standardised soil moisture anomalies in the weeks leading up to the hot spells. Right-hand side panels (b,d,f,h) show the probability of long and short hot spells preceeded by 60-day and 20-day periods of dry soils as well as during the first 4 days of the events.

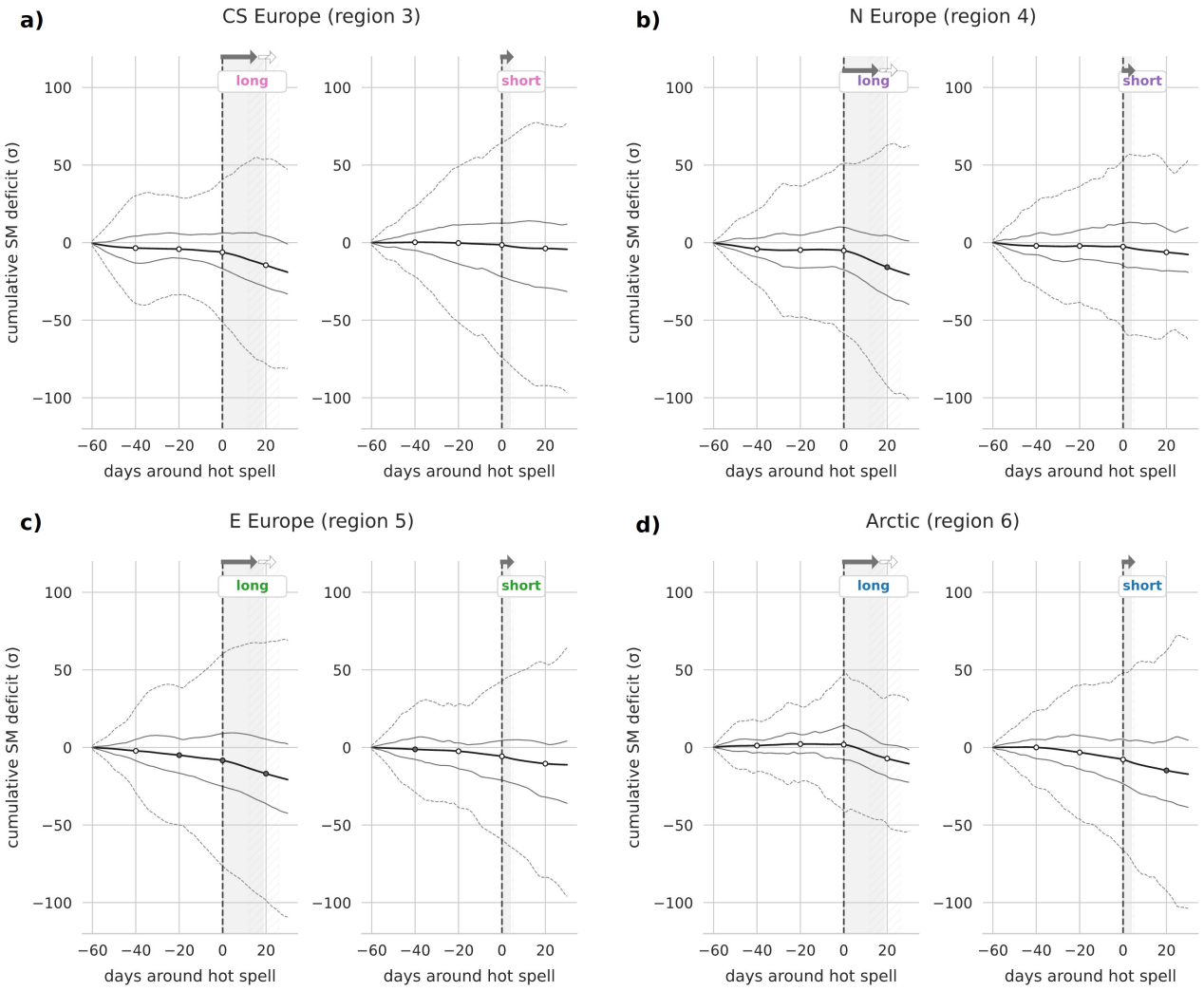

Figure S15. SM analysis analogous to Fig. 4 in Sect. 3.2 from the main manuscript for CS Europe (region 3, a), N Europe (region 4, b), E Europe (region 5, c) and Arctic/NW Russia (region 6, d). It depicts cumulative SM curves calculated with standardised anomalies of volumetric soil water content around long and short spells. The x axis ranges from 60 days before to 30 days after hot spell onset. The curves denote the mean (thick black), the 1st and 2nd quartiles (gray), and the lower and upper whisker limit values (thin gray dashed). Filled circles indicate timesteps at which the mean value is significantly different from zero at the 95% level.

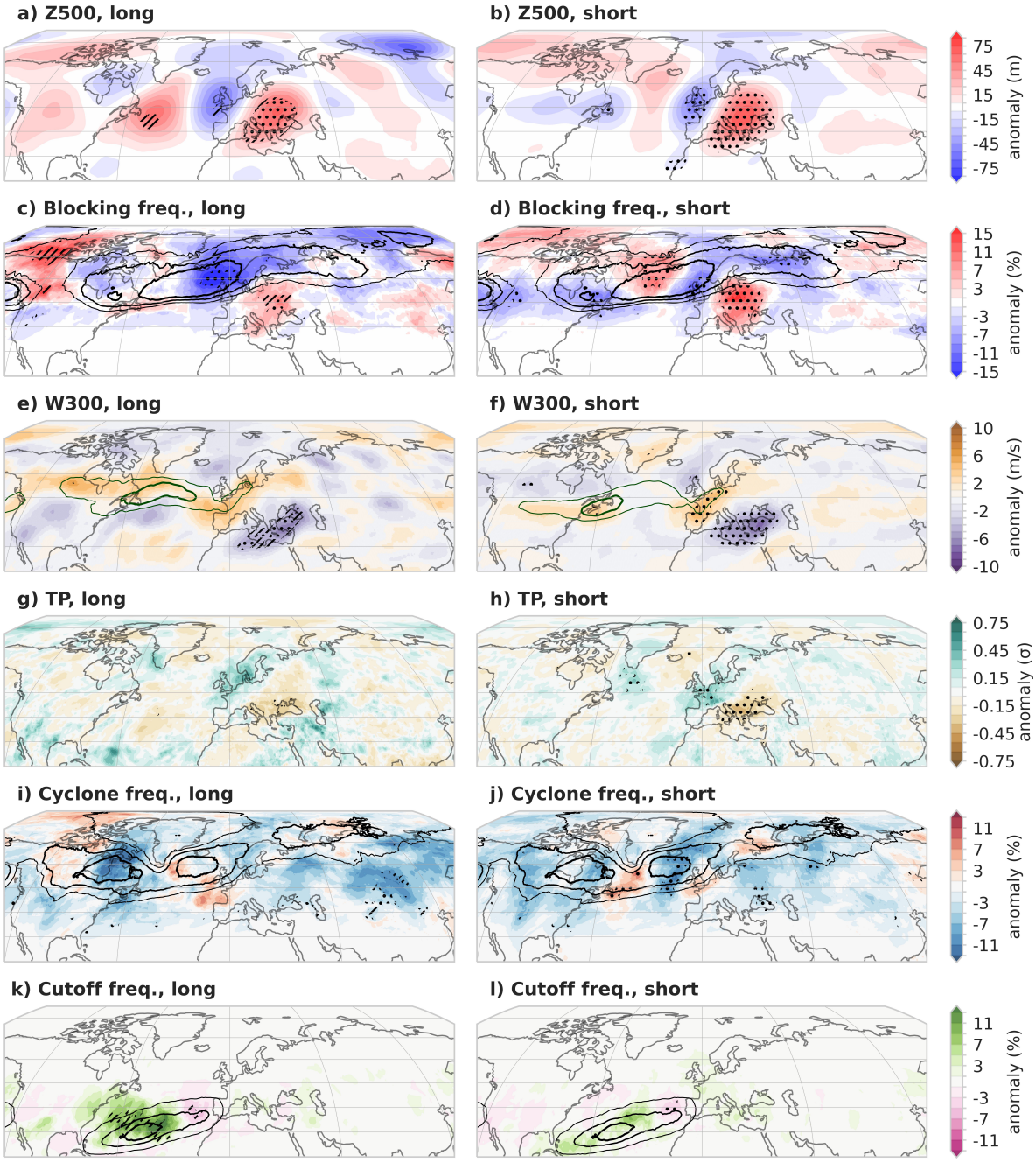

Figure S16. Composite fields for long (left column) and short (right column) hot spells in CS Europe (**region 3**). The variables Z500, W300, and TP represent geopotential height at 500 hPa, wind speed at 300 hPa, and total precipitation, respectively. Dashed (dotted) hatching in the long spells denotes areas where the anomalies are significant at a 95% confidence level in over one-third (two-thirds) of the 100 subsampled composites; dotted hatching in the short spell composites denotes significant anomalies at the 95% level.

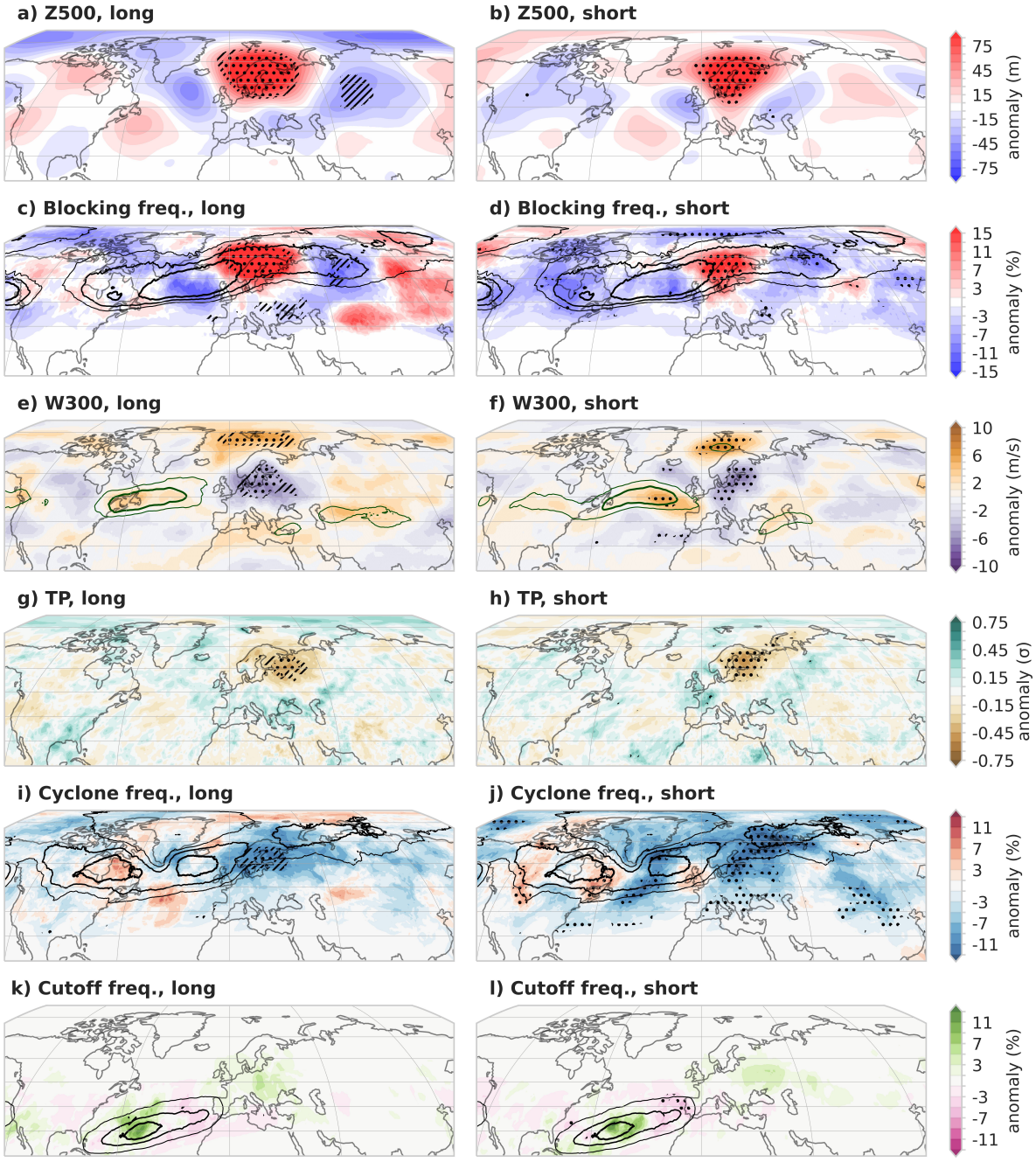

Figure S17. Composite fields for long (left column) and short (right column) hot spells in N Europe (**region 4**). The variables Z500, W300, and TP represent geopotential height at 500 hPa, wind speed at 300 hPa, and total precipitation, respectively. Dashed (dotted) hatching in the long spells denotes areas where the anomalies are significant at a 95% confidence level in over one-third (two-thirds) of the 100 subsampled composites; dotted hatching in the short spell composites denotes significant anomalies at the 95% level.

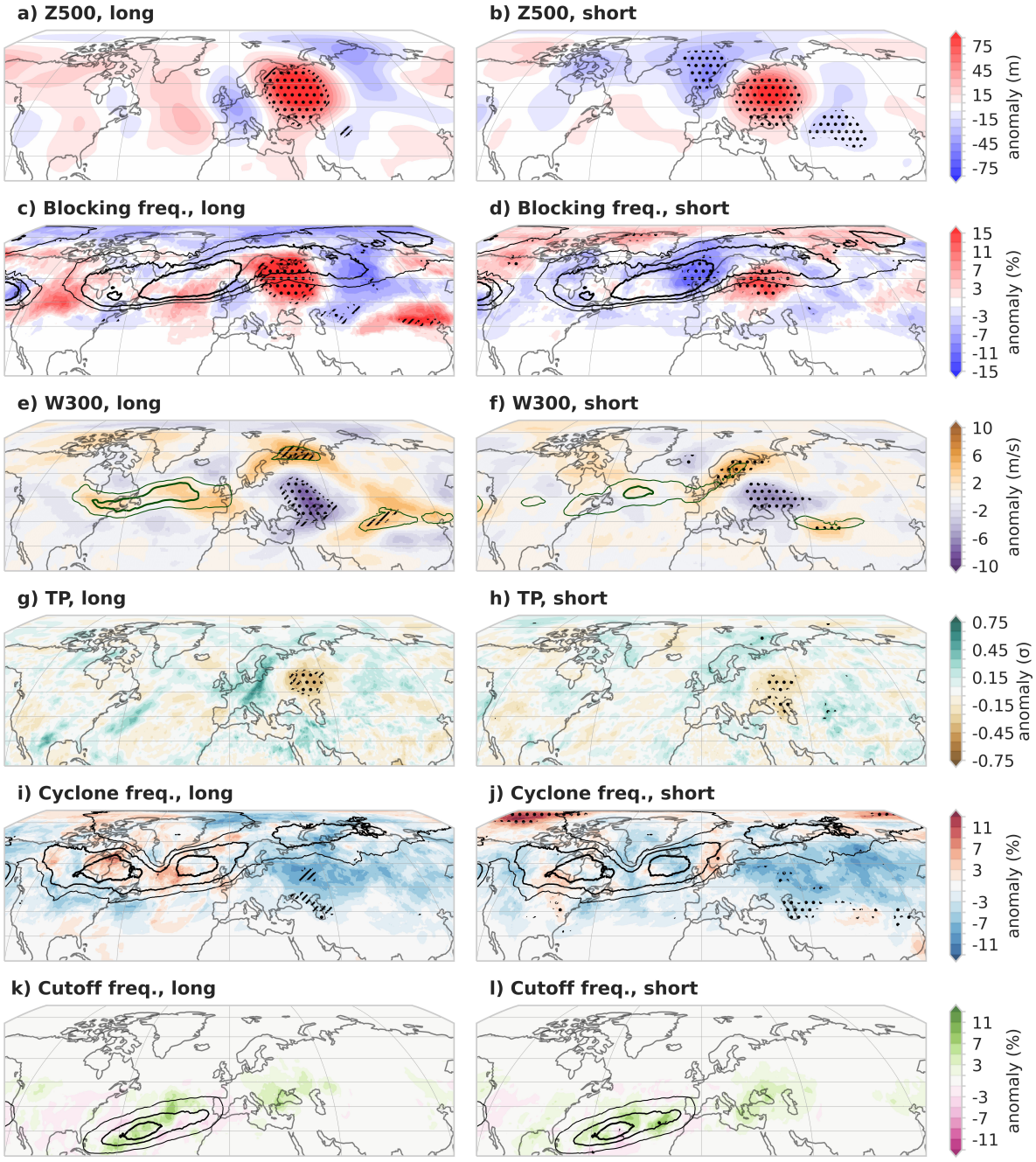

Figure S18. Composite fields for long (left column) and short (right column) hot spells in E Europe (**region 5**). The variables Z500, W300, and TP represent geopotential height at 500 hPa, wind speed at 300 hPa, and total precipitation, respectively. Dashed (dotted) hatching in the long spells denotes areas where the anomalies are significant at a 95% confidence level in over one-third (two-thirds) of the 100 subsampled composites; dotted hatching in the short spell composites denotes significant anomalies at the 95% level.

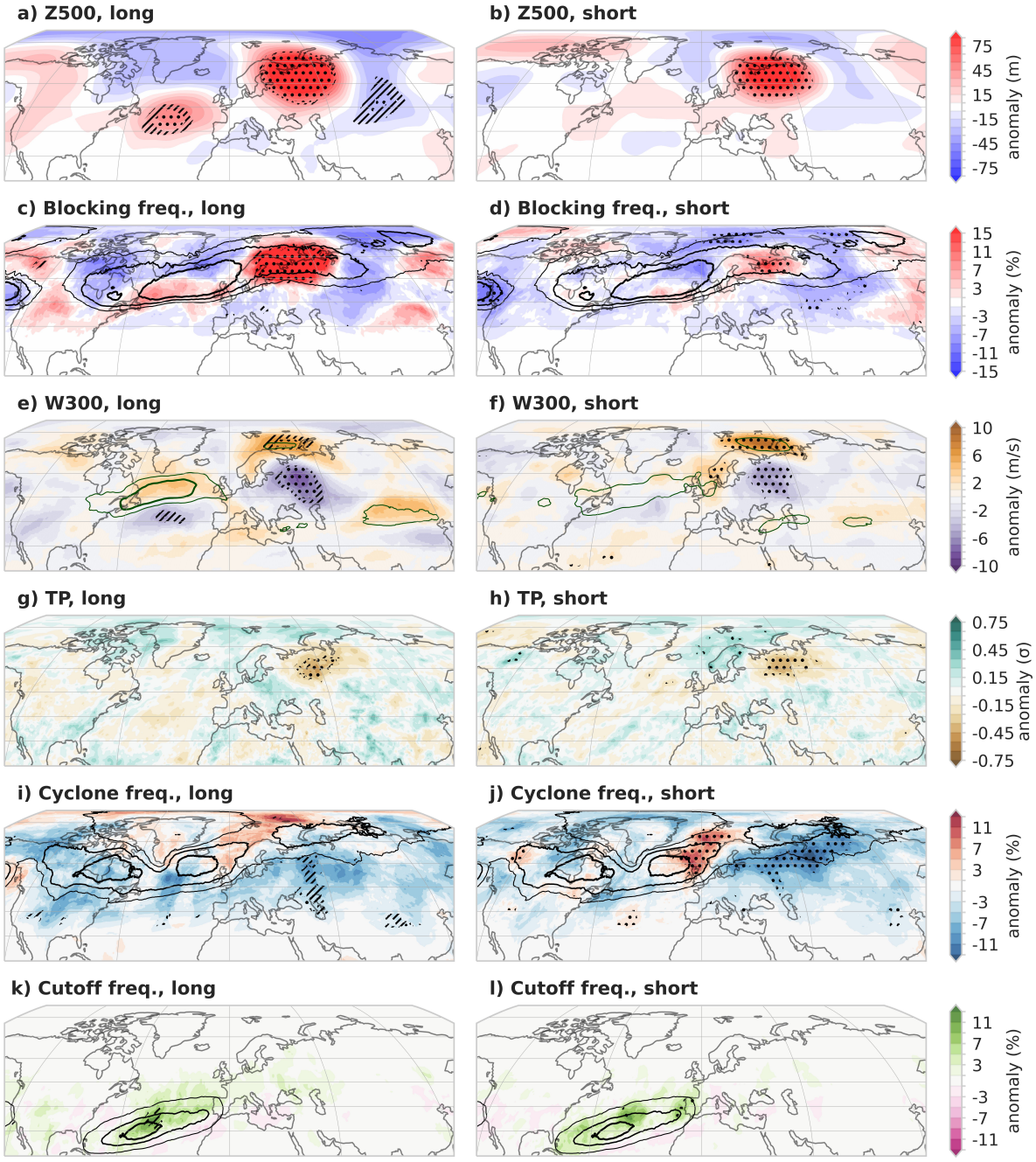

Figure S19. Composite fields for long (left column) and short (right column) hot spells in Arctice/NW Russia (**region 6**). The variables Z500, W300, and TP represent geopotential height at 500 hPa, wind speed at 300 hPa, and total precipitation, respectively. Dashed (dotted) hatching in the long spells denotes areas where the anomalies are significant at a 95% confidence level in over one-third (two-thirds) of the 100 subsampled composites; dotted hatching in the short spell composites denotes significant anomalies at the 95% level.

5 Full-spell composites of persistent hot spells

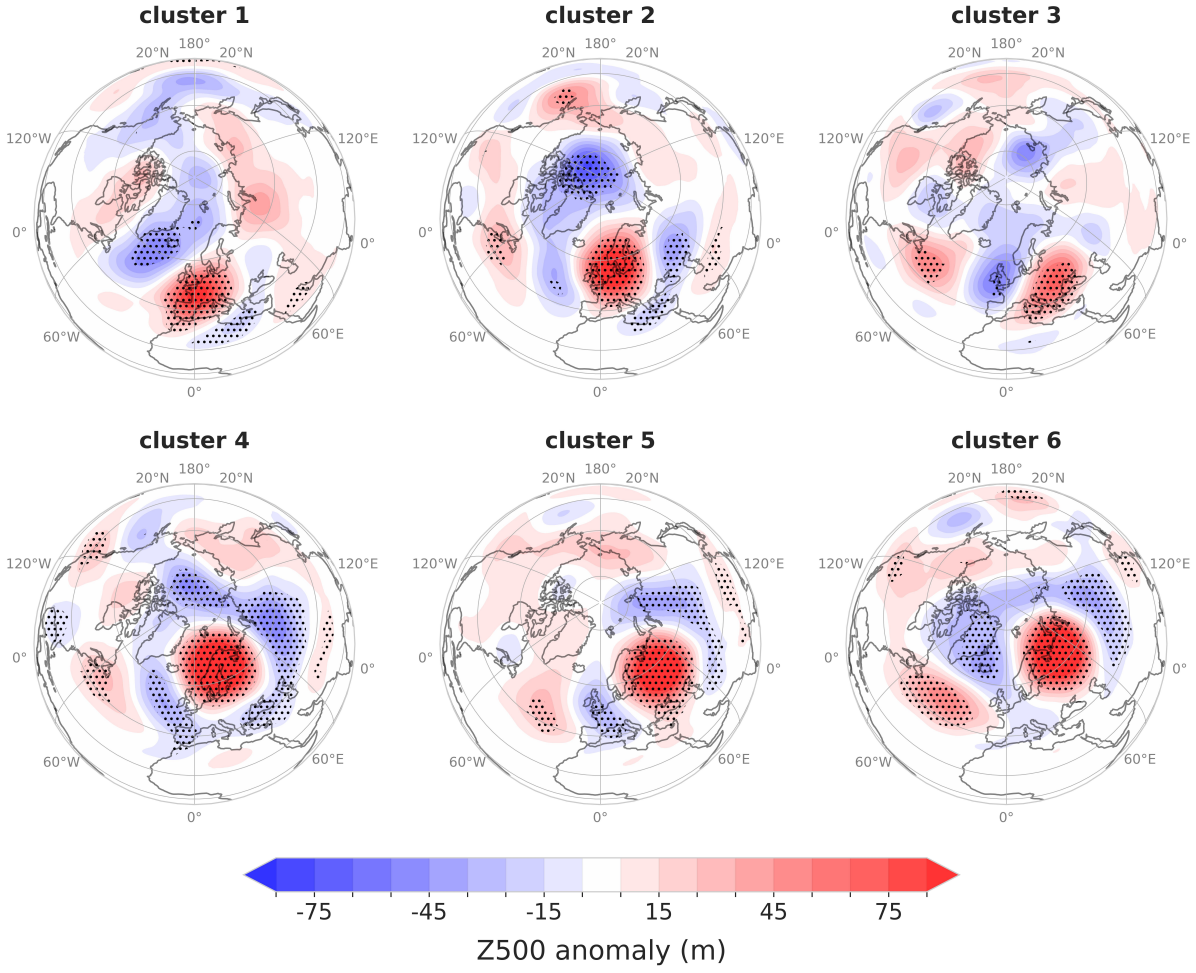

Figure S20. Spatial composites of geopotential height anomalies at 500 hPa for all six regions introduced in the study. Clusters refer to the following regions: 1) SW Europe, 2) W Europe, 3) CS Europe, 4) N Europe, 5) E Europe, and 6) Arctic/NW Russia.

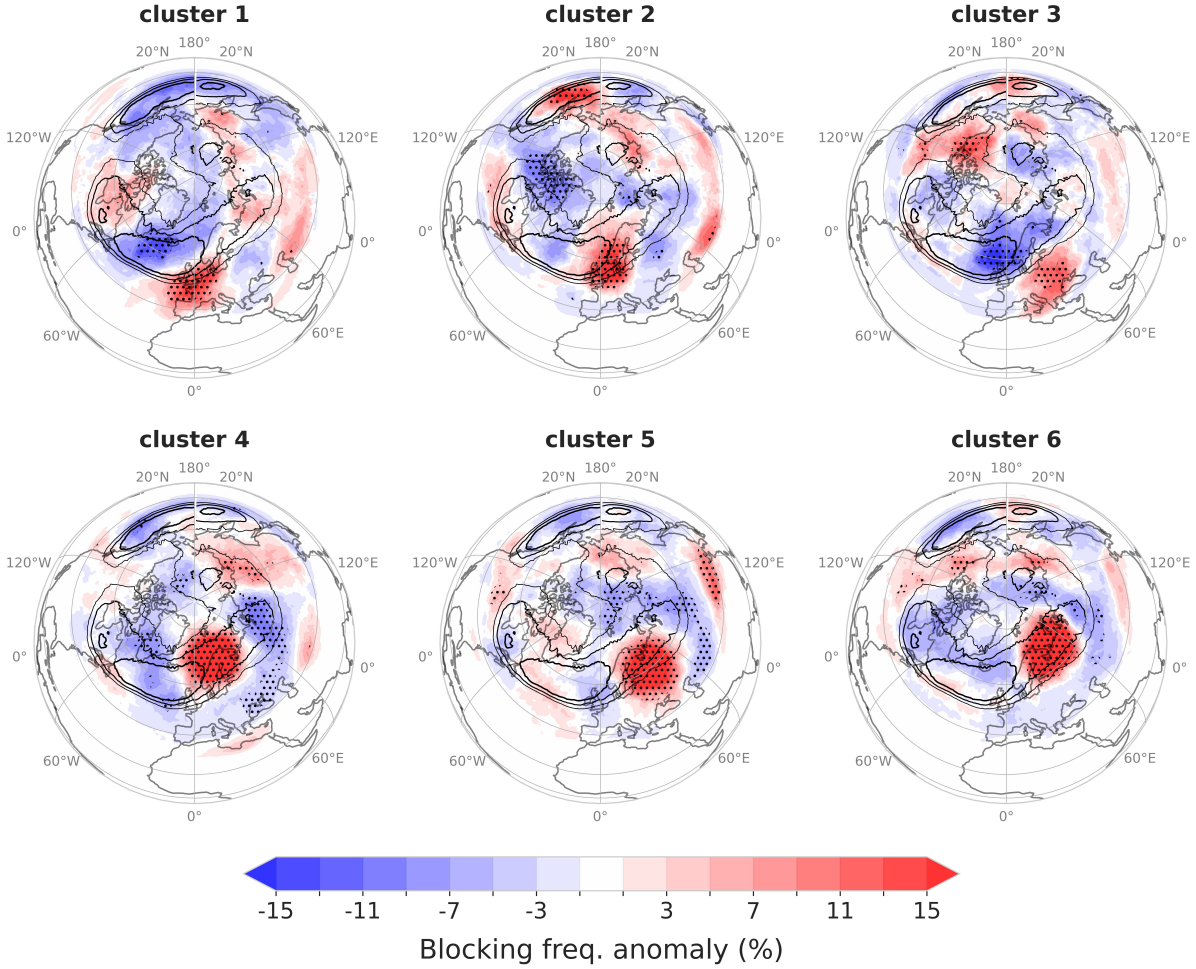

Figure S21. Spatial composites of blocking frequency anomalies for all six regions introduced in the study. Clusters refer to the following regions: 1) SW Europe, 2) W Europe, 3) CS Europe, 4) N Europe, 5) E Europe, and 6) Arctic/NW Russia. Black contours denote the mean MJJAS frequency at 10, 12, and 14%.

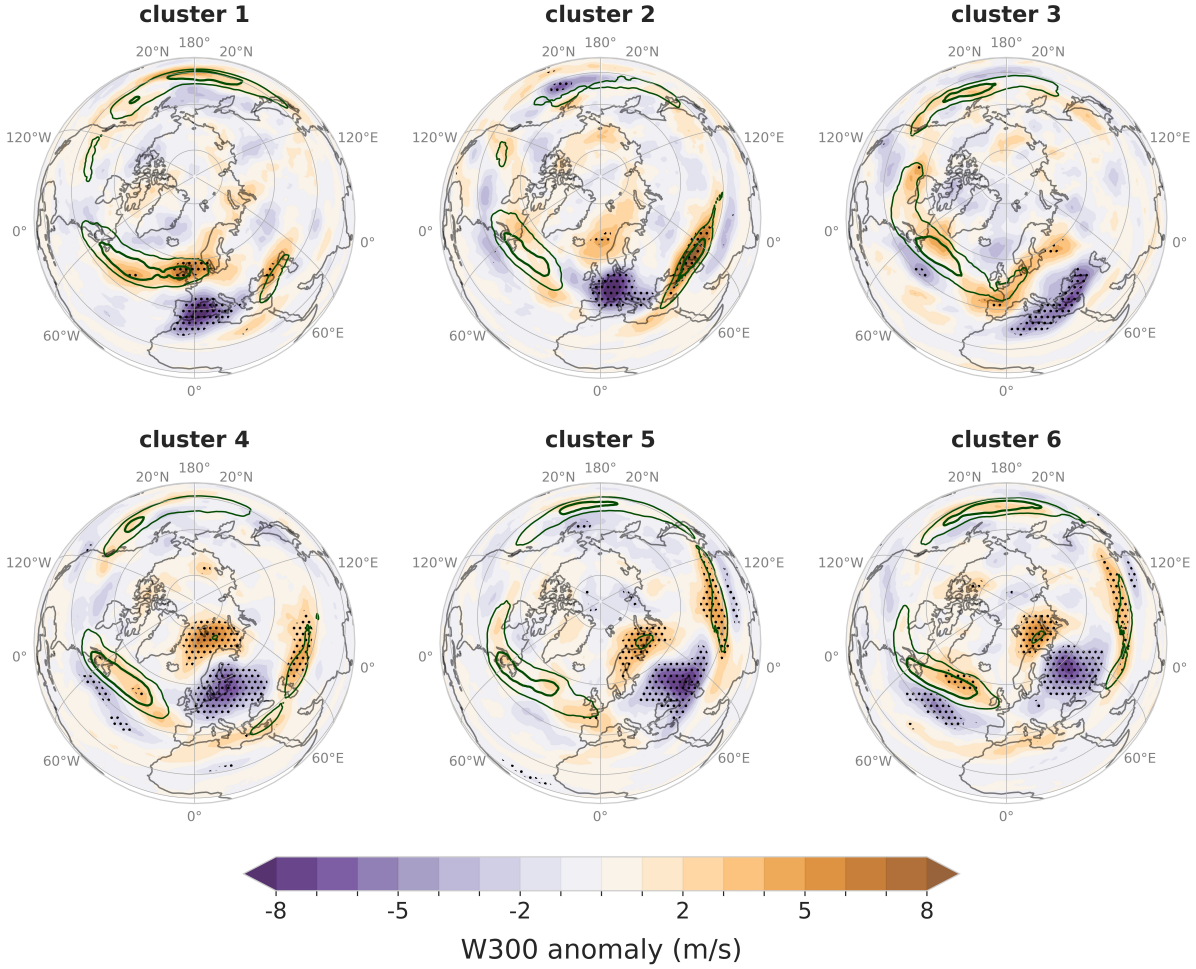

Figure S22. Spatial composites of wind speed anomalies at 300 hPa for all six regions introduced in the study. Clusters refer to the following regions: 1) SW Europe, 2) W Europe, 3) CS Europe, 4) N Europe, 5) E Europe, and 6) Arctic/NW Russia. Green contours show the corresponding absolute wind field to the anomalies (24 and 28 m/s).

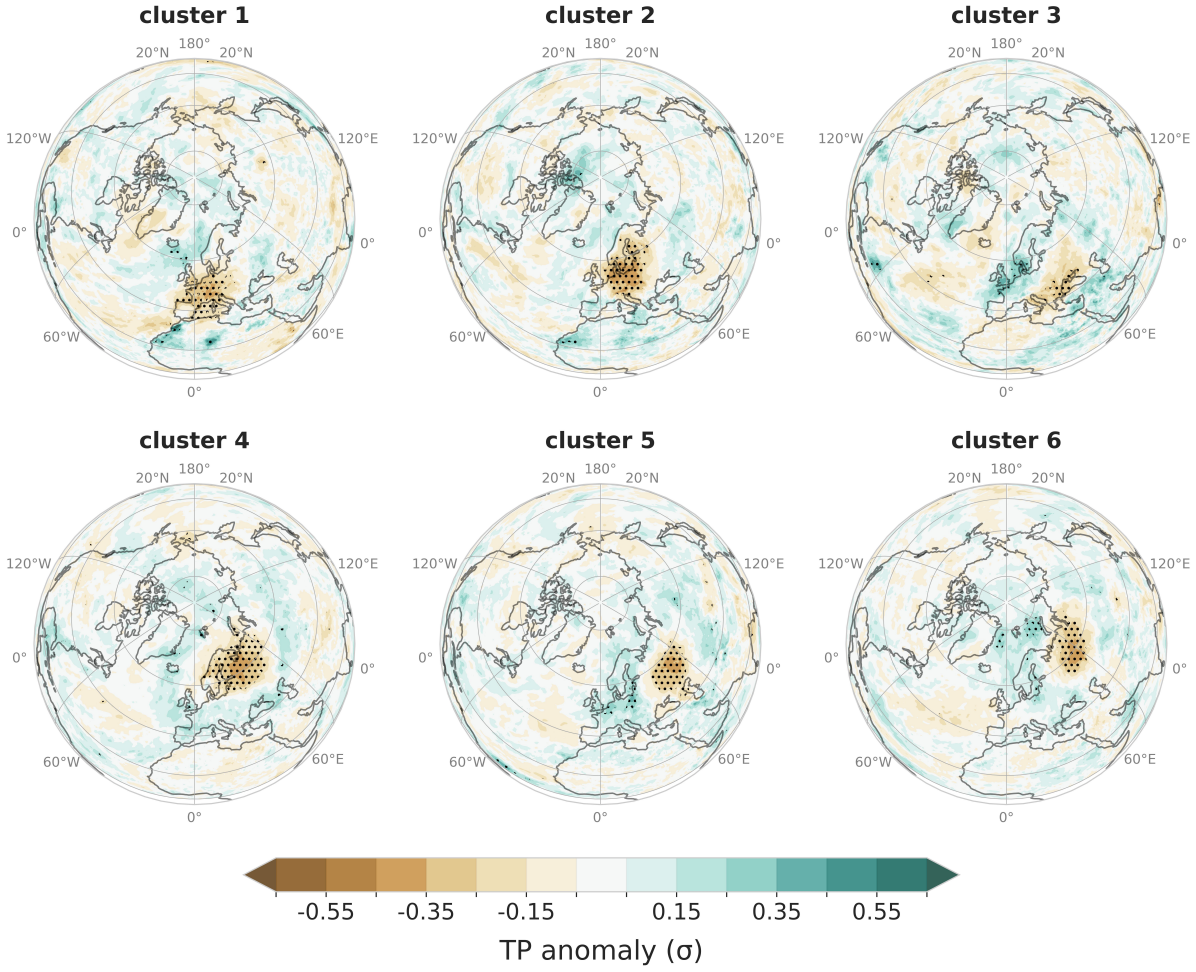

Figure S23. Spatial composites of total precipitation anomalies for all six regions introduced in the study. Clusters refer to the following regions: 1) SW Europe, 2) W Europe, 3) CS Europe, 4) N Europe, 5) E Europe, and 6) Arctic/NW Russia.

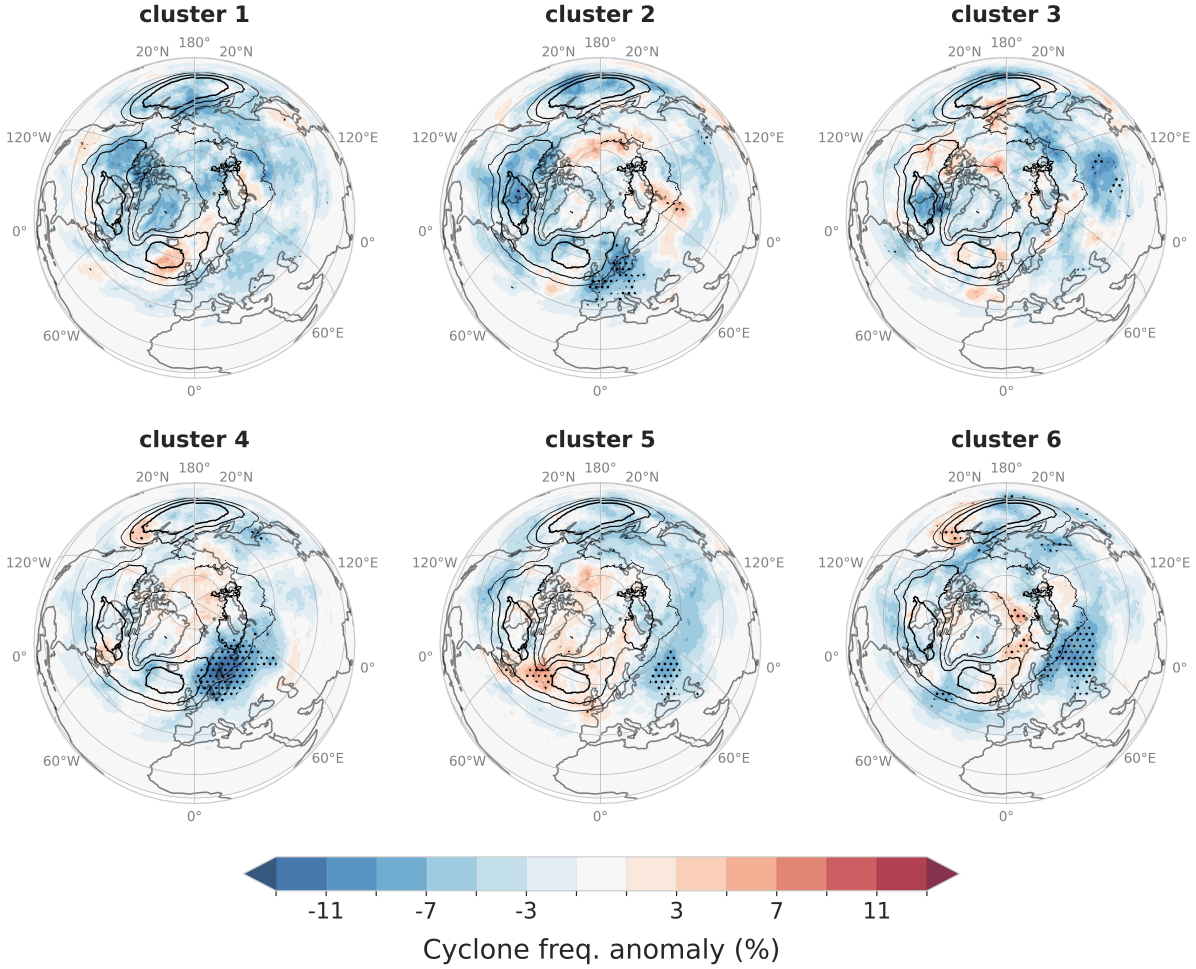

Figure S24. Spatial composites of cyclone frequency anomalies for all six regions introduced in the study. Clusters refer to the following regions: 1) SW Europe, 2) W Europe, 3) CS Europe, 4) N Europe, 5) E Europe, and 6) Arctic/NW Russia. Black contours denote the mean MJJAS frequency at 4, 7.5, and 10%.

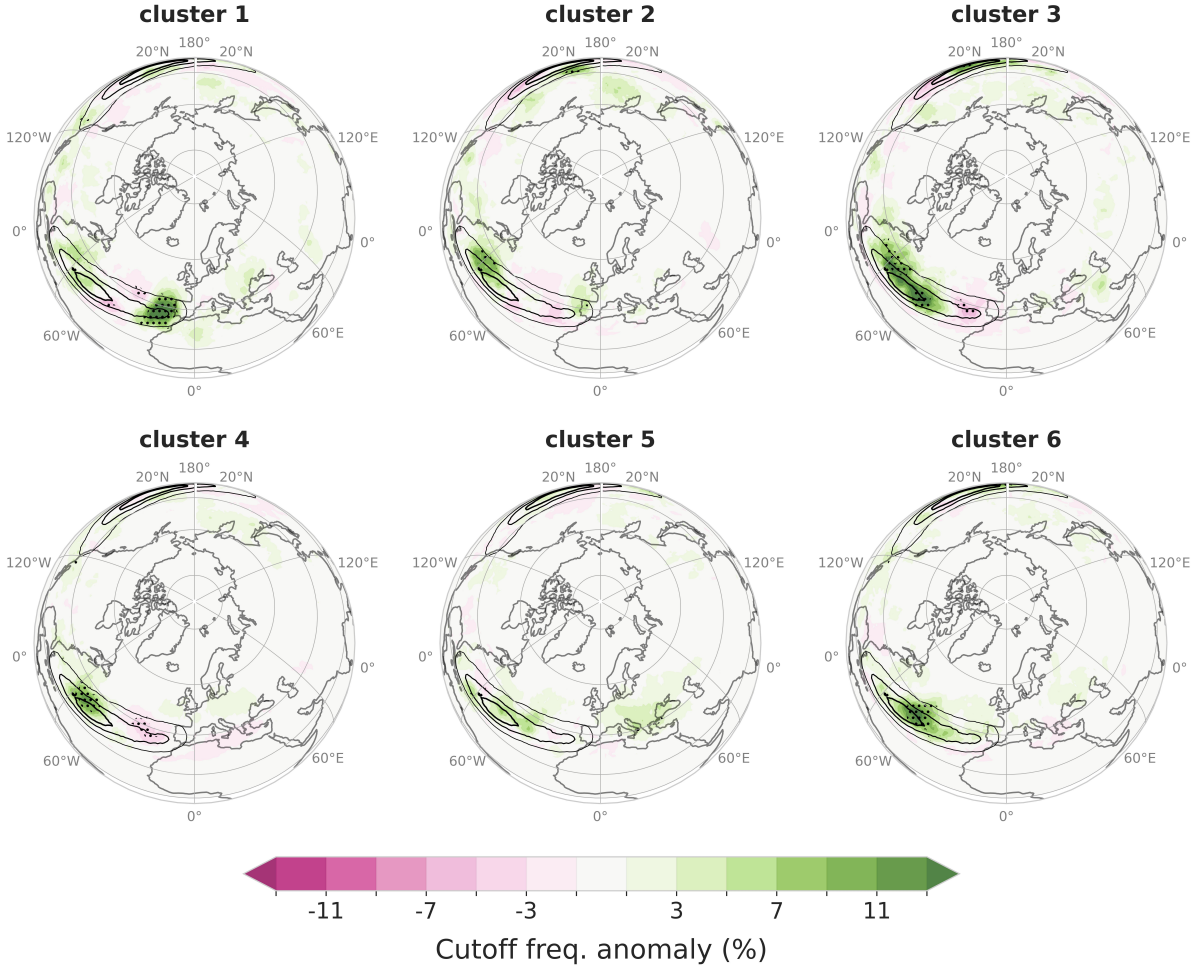

Figure S25. Spatial composites of stratospheric-PV cutoff frequency anomalies for all six regions introduced in the study. Clusters refer to the following regions: 1) SW Europe, 2) W Europe, 3) CS Europe, 4) N Europe, 5) E Europe, and 6) Arctic/NW Russia. Black contours denote the mean MJJAS frequency at 14, 17, and 20%.